DOCUMENT RESUME

ED 294 909 TM 011 616

AUTHOR Shayne, Vivian T.
TITLE Content Specification for Test Development.
PUB DATE [87]
NOTE 24p.; Attachments are marginally legible. Best copy available.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Content Analysis; *Item Analysis; Models; *Pretests Posttests; *Test Construction; Test Interpretation; Test Items; Test Reliability; Trainers; Welfare Services
IDENTIFIERS Experts; Income Maintenance; *Partitioning Procedures

ABSTRACT

A target partition analysis (TPA) was used to help seven subject matter experts, with little expertise in testing, evaluate relationships among test items. The subjects were experts in Income Maintenance in the field of public assistance work; they were trainers affiliated with the Professional Development Program's Income Maintenance Training Institute. Content specification of test items was achieved by comparing and contrasting individual expert's categorizations of test items. Expert models were evaluated, using the TPA, in terms of category clarity, similarity to other expert models, response style, and model elegance and parsimony. These analyses were also conducted to facilitate the development of internally consistent "testlets," which could be used to better evaluate pre- to post-test change. Results suggest that the TPA is useful in helping subject matter experts in test development efforts. (Author/TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CONTENT SPECIFICATION FOR TEST DEVELOPMENT

Vivian T. Shayne[1]
Rockefeller College
University at Albany
State University of New York

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
VIVIAN T. SHAYNE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED294909

A Target Partition Analysis was used to help subject matter experts, with little expertise in testing, evaluate relationships among test items. Content specification of test items was achieved by comparing and contrasting individual experts' categorizations of test items. Expert models were evaluated, using the TPA, in terms of category clarity, similarity to other expert models, response style, and model elegance and parsimony. These analyses were also conducted to facilitate development of internally consistent "testlets" which could be used to better evaluate pre- to posttest change. Results obtained suggest that the TPA is useful in helping subject matter experts in test development efforts.

PROBLEMS INHERENT TO CONTENT SPECIFICATION

Substantial progress in differentiating examinee ability and predicting relative standing has been achieved. However, the quantitative test models developed to facilitate assessment of individuals have not proven very useful in evaluating instruction. Over the years, traditional models of test development have been criticized for failure to reflect instructional course content (Bormuth, 1970; Hively, 1974; Lovitt, 1986) and to specify what the learner must know to perform successfully (Scandura, 1977). Translating subject matter into test items remains problematic (Bejar, 1985, 1987; Embretson, 1985).

---

[1] The author is grateful for the invaluable assistance and input of Robert Pruzek and Dale Bryant.

TM 011 616

Because of its success in estimating measurement error, facilitating analysis of item bias, and equating; Item Response Theory continues to dominate test development efforts. However tests which violate assumptions of unidimensionality continue to be problematic for IRT. Nonetheless, test developers still create items which measure more than one ability. For educational activities designed to foster integration of different types of knowledge or skill, special consideration in assessment and evaluation may be warranted (Linn, 1986; Reckase, 1985). Subjects which are inherently multidimensional may not be assessed adequately by instruments designed to conform to restrictive requirements of IRT. Rigid adherence to IRT can also obscure understanding of processes mediating learning, particularly when different examinees employ different strategies to solve the same problem (Mislevy and Verholst, 1986).

As a result, psychometricians have attempted to extend IRT, or develop suitable alternatives, in order to better address item content rather than characteristics of individual examinees. Increased consideration is being given to cognitive processes mediating learning of specific knowledge domains of knowledge.

More tests are being developed to identify examinees' problem-solving strategies. Tests designed to discriminate among student errors have been used to diagnose misconceptions in learning (Brown and Burton, 1987; Cronbach and Snow, 1977; Taksuoka, 1983). Test items have been designed to investigate complex respondent capabilities (Scheuneman and Steinhaus, 1987).

Differential patterns of correct response probabilities have been used to make inferences about examinee response strategies.

Test items are specially constructed to enhance or deter performance depending upon examinees' choice of problem-solving strategy. Mislevy and Verholst (1986) address IRT mixture models in this context.

But such models all presuppose some description of item dimensionality or course structure. Descriptions have been guided by cognitive process models (Sternberg, 1985; Wilson, 1984, 1985) or, depending upon subject matter, more informally.

Unfortunately, the expertise of most test developers is confined to some relevant subject matter (e.g. English or history), or curriculum development. Moreover, much of this expertise is intuitive. The significance of psychological theories, or research methods, relevant to test item construction is limited, and rarely guided by empirical research (Embretson, 1985). As a result, test development, particularly for unusual topics is not very systematic.

Although processes of test item development and selection are considered logical, or deductive (Guttman, 1941) in reality, basic elements of a particular content domain often elude identification. Different though equally qualified persons are likely to disagree in their descriptions of item difficulty, discrimination or other content features (Bejar, 1983).

Because test development efforts are largely contingent upon some prior specification of subject matter organization or knowledge structures, judgmental methods to help articulate and evaluate working knowledge of even experienced test developers are needed (Olson et al, 1987).

METHODOLOGY FOR INVESTIGATING CONTENT SPECIFIATION

Partition analyses have been used previously in this context. Essentially, these procedures require curriculum experts to sort test items into homogeneous categories representing relevant content characteristics. Although categories may be labeled or defined differently by each expert sorter, some consensus is likely to be demonstrated among experts in their grouping of certain items together, while keeping others separate.

Latent partition analysis (Wiley, 1967) was specifically developed to derive homogeneous item subsets concerning more specific aspects of a given instructional domain. Hartke (1978) used LPA in a logical judgment process to test for conceptually homogeneous population. However, LPA, like common factor analysis, is complicated by certain indeterminancy problems and may be regarded as "blind" to the extent that prior beliefs do not guide analysis (Pruzek, Stegman, and Pfeiffer, 1972).

In contrast, target partitioning analysis (TPA)(Pfeiffer, Pruzek, and Sherry, 1975; Pruzek, Stegman, and Pfeiffer, 1972) was specially designed to elicit prior hypotheses, or beliefs, about structural relationships among items, and may be regarded as somewhat Bayesian in this respect.

The TPA describes the relationship between test items through a series of binary joint occurrence matrices. The matrix describing item categories for a target (A_t) is compared to joint occurrences of items for other sorters (A_s). The proportion of difference in predicting joint occurrences for sorters given the target partition matrix is described by the q_{st} trace statistic.

Accordingly:

$$q_{st} = \sum (A_S (I,J) - A_T (I,J))^2 / (N*(N-1))/2$$

where

A (i, j) = 1, if items i and j occur together
= 0 otherwise

When all sorters (including target) categorize items identically $q_{st}=0$.[2] Alternatively, $q_{st}=1$ in the absence of agreement between sorters and target.

In addition to the joint proportion matrices describing emergent relationships between the target's categorization model to others' categorizations of test items, the TPA program[3] also produces confusion matrices which focus attention on the relationships between item categories. Accordingly, diagonal elements of the confusion matrix describe the degree of overlap between target and sorters in grouping a given set of items together, while off-diagonal elements describe the degree of relation for that set of items to the target's other categories. (see Tables 1b, 2b, 3b and 4b). Close inspection of the confusion matrix can be very informative. The location of a high off-diagonal value pinpoints the source of confusion between categories. For example, the high off diagonal value in the first column and seventh row in the confusion matrix of table 1B

---

[2] This lower value is difficult to attain because achievement is contingent upon having identical category frequencies for an identical number of categories in comparison and target matrices.

[3] The target partition analysis program was initially developed by Robert Pfeiffer and Robert Pruzek at the University at Albany, State University of New York.

C

(on page 18) suggests other sorters were inclined to group the single item constituting this particular target's first category, with items the target includes in category seven. In contrast, confusion between items in the fourth or fifth categories appeared minimal in this expert's model (See intersection between the fourth column and the fifth row, on Table 1B on page 18).

The program also has the capability to derive modal partitions or hierarchies for items. This can be used to further refine and better evaluate models generated by expert sorters.

The TPA has been used to compare item sorts of students with varying levels of sophistication to subject matter experts. As expected, more knowledgeable students appeared closest to the target's sort (Ambrosino, McMorris, and Naval, 1979).

Because the TPA is specially designed to facilitate the evaluation of prior belief, this method is seems promising as a guide for subject matter experts with little or no expertise in test development.

AN EMPIRICAL STUDY OF TARGET PARTITION ANALYSIS USED FOR CONTENT SPECIFICATION

In an effort to explore the usefulness of TPA in helping subject matter experts articulate their beliefs about how test items relate to each other, attention was focused on a group of highly specialized trainers affiliated with the Professional Development Program's Income Maintenance Training Institute. This program was designed to prepare participants to assume the role of public assistance workers responsible for processing welfare clients. In compliance with New York State Department of Social

Service evaluation guidelines, administration of a pre-post test is mandatory. Although trainers have expertise in Income Maintenance, they lack expertise in test development. Item analyses routinely conducted on test results have proven of little use in helping trainers develop new items, or replace old items which have become obsolete due to changes in federal or state regulations.

The seven trainers categorized fifty items from the test used to evaluate delivered training. The TPA was used to compare and contrast different trainers' knowledge models vis a vis the group as a whole. Expert models were compared and contrasted in terms of response style, category clarity, similarity to other experts' models, and overall model elegance and parsimony. Profiles describing the relationship of each trainer's model to other trainer categorizations were generated. Excerpts from these are appended (see Tables 1-4).

The TPA was also used to evaluate individual test items. Items most consistently grouped together were presumed to be related more closely to each other. Such items were distinguished from items which appeared in many different categories; and from items which consistently failed to relate well to any given item cluster.

The efficacy of each trainer's knowledge model, vis a vis the group, as a whole was evaluated by examining average q_{st} scores to see which were lowest and highest. Because of its sensitivity and generality, the q_{st} statistic is considered to be a particularly appropriate indicator of how well items (or objects)

categorized together really belong together (Boorman and Arabie, 1972).

But the q_{st} is inherently biased toward targets who are inclined to use many small categories to describe the relationship between items. Such "splitters" may be distinguished from "lumper" targets who generate comparatively few categories, each containing many items.

Even during sorting, these response styles were evident. Splitters expressed visible signs of discomfort when forced to consolidate many small piles into ten categories (the maximum number permitted). Alternatively, lumpers appeared more intent on reviewing, rechecking, or reorganizing categories which were fewer in number but larger in size.

Results obtained were difficult to evaluate in terms of average q_{st} because differences were often small (see Tables 1A-4A). The trainer with the largest q_{st} exhibited a "lumper" response style (see Table 4). Because this individual was regarded as one of the most competent and experienced trainers, attention was focused on other parameters in order to better evaluate the significance of discrepancy between this individual and his peers. Individual q_{st} scores between target and sorter were also examined to better evaluate the magnitude of similarity or difference. The trainer who attained the lowest average q_{st} was more often similar to other sorters while the trainer with the highest q_{st} was most often most discrepant (see Tables 1A-4A).

The pattern of results obtained was also carefully scrutinized in terms of category clarity as shown by confusion

matrices. Inspection of such matrices shows the extent to which items included in a given category were more similar to each other than to items included in other categories.

An index of category clarity was developed to further facilitate such description. This measure was also developed in view of the q_{st} splitter bias.

This index was created to take into account category size as well as number of categories. Essentially, this index is a ratio of the weighted average of diagonal values in the confusion matrix to the weighted average of off-diagonal values. Accordingly:

$$CCI = \frac{\sum_{i}^{k} \frac{n_i}{N} \frac{C_{jj}}{k}}{\sum_{i=1}^{k} \sum_{j=1}^{i} \sqrt{\frac{n_i\, n_j}{N}} \frac{C_{ij}}{k(k-1)/2}}$$

CCI = Category Clarity Index
n_i = number of items per partition
k = number of categories
C_{jj} = diagonal values
C_{ji} = off-diagonal values
N = total number of items

RESULTS OBTAINED

As expected, the trainer with the lowest q_{st} (see Table 2) obtained a comparatively high score on the Category Clarity Index[4], while the trainer with the highest q_{st} obtained the lowest score on this index (see Table 4). Category clarity appeared less susceptible to response style, and at least for

this sample, appeared to be more sensitive to individual variation (see Table 5).

Trainers were asked to describe which categories were easiest to generate and which were most difficult. But experience of difficulty had little relation to category clarity. Conversely, some categories described as easy to generate were not consensually validated by their peers.

This is not surprising. In the past, subject matter experts have not proven good judges of item difficulty or discrimination (Bejar, 1983). Categorization of even basic-level objects which can be readily named and recalled are often less accessible psychologically. Moreover, categories which lack firm boundaries may still be seen similarly by others (Gardner, 1985).

Responses to items were much easier to interpret. A number of items which did not fit any model in were readily identified. Some were be targeted for elimination. Other items, or item subsets, which were subsumed in different categories by different sorters were evaluated in terms of their prospective multidimensionality. Because the subject matter at hand is largely multidimensional, consideration of creating "testlets" to better evaluate pre- to post test has been entertained.

Partitioning tests into ordered "testlets" appears to obviate many difficulties associated with conventional adherence to total or percent correct score, including failure to represent initial status gain, floor to ceiling effects and the failure to

---

[4] Failure to obtain the highest score reflects this individual's creation of a miscellaneous category for one item, generally regarded by other sorters as not fitting well with any category.

capture any sense of subject organization. Subtests or testlets, parcels have become increasingly popular among psychometricians dissatisfied with the conventional emphasis accorded to total, or percent correct scores (Cook et al, 1985; Rock and Pollack-Ohls, 1987). Future study will focus on evaluation of pre- to post test change in terms of testlets derived.

SUMMARY

Knowledge is not fixed and immutable but subject to question. Therefore theories describing the structure of any given domain are subject to change. Cognitive "mapping" of specialized, comparatively esoteric domains is generally relegated to subject matter experts responsible for instruction and curriculum development. Subject matter experts may experience difficulty and frustration in constructing test items which adequately represent knowledge structure.

Although test development is often described as a linear process (Hively, 1974; Roid and Haladyna, 1980), in actuality, because items are used to both measure and define constructs, this process is more iterative (Messick, 1979; Pruzek and Lehere, unpublished). Therefore, it is especially important to establish a context for ongoing development efforts.

The role of subject matter experts responsible for instruction has been neglected by researchers of test development. However, even when there is a strong theoretical foundation describing organization of course content, it is still important to consider instructors conceptualizations. Because

cognitive processes mediating learning for various knowledge domains is often poorly understood, it seems especially important to actively involve subject matter experts in test development and design. When tests measure things that are not being taught, or poorly measure things that are being taught, they must fail to achieve instructional objectives. The lack of a responsive testing policy may be debilitating and demoralizing and can even discourage instructors from good teaching. This is especially problematic when teaching and assessment activities are not well integrated (Lovitt, 1986).

Testing plays a critical role in curriculum development and planning. But success in this direction appears to warrant collective effort on the part of highly specialized experts who are often unfamiliar with principles of test development, psychology, or learning theory.

Methods like the TPA are especially promising in this respect. The TPA is suitable for small groups of experts working closely together. The procedure has also proven useful in investigating parallels in thinking between expert instructor and students. Differences between students at various levels (e.g. advanced, intermediate and basic) can also be evaluated.

Unfortunately, although this procedure was developed some time ago, it has not received very much attention. The usefulness of statistics like the q_{st} or the category clarity index requires further investigation.

The intent in this study was to stimulate experts into thinking more systematically about how they collectively organize subject matter. Results obtained suggest that subject matter

experts cannot anticipate how they compare to peers in organizing constructs.

The TPA appears useful in helping curriculum experts articulate their unique points of view regarding subject matter. It also provided trainers with feedback regarding their model's acceptance by peers. Of course, models which are not generally accepted by others may still be better than models which are accepted. But the TPA can inform experts why their models are problematic for others.

Trainers also obtained a better sense of how their response styles influence their thinking. This is valuable for subject matter experts lacking confidence in their test development skills, and provides insights regarding other aspects of instruction as well. One experienced trainer, a lumper, described how she had been accused of using examples other trainers reserved for discussion of specific topics. For her, the topics were more interrelated. Knowledge of her response style provided a better sense of perspective.

Items which failed to relate to critical constructs were targeted for elimination. Trainers also acquired a better understanding of items which may represent multiple overlapping constructs or dimensions. Perhaps most important of all, trainers now have shared understanding of context for future discussion. For this small group of subject matter experts, these outcomes represent significant milestones.

Increasingly, knowledge and understanding are developed and enhanced by computer simulations, heuristic demonstrations, and

reliance on complex equipment and instrumentation. Advances in knowledge are less dependent upon the extended work of single individuals, and increasingly dependent upon the enterprise of complex communities and large institutions (Bohm and Peat, 1987; Galison, 1987). Ideas generated must be understood in terms of an intricate evolving, shared context of communication.

Thus, it is imperative to develop better mechanisms for understanding differences in expert belief systems in order for educators to optimally synchronize themselves with each other and their students. Such mechanisms are a necessary antecedent to test development and instruction, and merit continuing attention.

REFERENCES

Ambrosino, R.J., McMorris, R.F., and Naval, L.K. (1979). Partitioning methods for detecting misconceptions of content and test items. Journal of Educational Measurement, 16(3), 187-195.

Arabie, P. and Boorman, S.A. (1973). Multidimensional sealing of measures of distance between partitions, Journal of Mathematical Psychology, 48-203.

Bejar, I.I. (1987, Apr. 21). Analogy items: Semantic relationship schemes and differential item functioning. Symposium conducted at the meeting of NCME, Washington, DC.

———. (1985a). Speculations on the future of test design. In S.E. Embretson (Ed.), Test Design: Contributions from Psychology, Education, and Psychometrics.

———. (1985b). Speculations on the psychometrics of test design. In S. E. Embretson (ed.), Test Design: Contributions from Psychology, Education and Psychometrics, pp. 279-292. New York: Academic.

———. (Summer 1983). Subject matter experts' assessment of item statistics. Applied Psychological Measurement, 7(3): 303-310.

Bohm, D., and Peat, F.D. (1987) Science Order and Creativity. Toronto: Bantam Books.

Bormuth, J.R. (1970). On the Theory of Achievement Test Items. Chicago: University of Chicago Press.

Bohrstedt, G.W. (1983). Measurement, in J. Wright and P. Rossi (eds.), Chap. 3, pp. 69-121 Handbook of Survey Research, New York: Academic Press.

Brown, J.S. and Burton, R.R. (1987). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.

Cook, L., Dorns, N., Eignor, D., and Petersen, N. (1985). An assessment of the relationship between the assumption of unidimensionality and the quality of IRT true-score equating (RR 85-30). Princeton, NJ: Educational Testing Service.

Cronbach, L.J., and Snow, R.E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. New York: Irvington.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Derry, S. J., and Hawkes, L. W. (1987). Training problem solving ability. Paper presented at the annual meeting of the American Educational Research Association, Washington, D. C.

Embretson, S.E. (1985a). Introduction to the problem of test design. In S.E. Embretson (Ed.), Test Design: Developments in Psychology and Psychometrics. Orlando: Academic Press.

———. (1985b). Multicomponent latent trait models for test design. In S.E. Embretson (Ed.), Test Design: Developments in Psychology and Psychometrics. Orlando: Academic Press.

———. (1985b). Component latent trait models of paragraph comprehension. Final report, Batelle contract 0855. Lawrence, KS: University of Kansas.

Faley, R.H., and Sandstrom, E. (1985). Content representativeness: An empirical method of evaluation. Journal of Applied Psychology, 70, 567-571.

Ford, J.K., and Wroten, S.P. (1984). Introducing new methods for conducting training evaluation and for linking training evaluation to program redesign. Personnel Psychology, 37, 651-665.

Gagne, E.D. (1986). The cognitive psychology of school learning. Boston: Little Brown.

Gagne, R.M. (1980). Learnable aspects of problem solving. Educational Psychologist, 15, 84-92.

Gagne, R.J. (1985). The conditions of learning (4th ed.). New York: Holt, Rinehart and Winston.

Galison, P. (1987) How Experiments End. Chicago: The University of Chicago Press.

Gardner, H. (1987). The Mind's New Science: a History of the Cognitive Revolution. New York: Basic Books, Inc.

Glaser, R. (1986). The integration of instruction and testing. In The Redesign of Testing for the 21st century. Proceedings of the 1985 ETS Invitational Conference.) Princeton, NJ: Educational Testing Service.

Glaser, R., and Lesgold, A. (1985). Congitive task analysis and the measurement of proficiency. Paper presented at the Buros-Nebraska Symposium on Measurement and Testing, Lincoln, Nebraska.

Guttman, L. (1969). Integration of test design and analysis. In Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service.

———. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (Ed.), Prediction of Personal Adjustment. New York Social Science Research Council.

Hartke, A.R. (1978) The use of latent partition analysis to identify homogeneity of item population. Journal of Educational Measurement, 15, 43-47.

Hively, W. (1974). Introduction to domain referenced testing. Educational Technology. 14: 5-9.

Linn, R.L. Educational testing and assessment: Research needs and policy issues. American Psychologist.

Lovitt, T.C., and Haring, N.G. (1979). Classroom applications of precision teaching. Seattle, WA: Special Child Publications.

Mislevy, R. (1985). Recent developments in the factor analysis of categorical variables (RR 85-24). Princeton, NJ: Educational Testing Service.

———. (1984) Estimating latent distributions. Psychometrika, 49, 359-381.

Mislevy, R.J., and Verholst, N. (1986). Modeling item responses when different subjects employ different solution strategies. Paper presented at the American Psychological Association.

Pellegrino, J.W., Mumaw, R.J., and Shute, V.J. (1985) Analysis of spatial aptitude and expertise. In S.E. Embretson (Ed.), Test Design: Developments in Psychology and Psychometrics. Orlando: Academic Press.

Pfeiffer, R.A., Pruzek, R.M., and Sherry, J.R. (1975) Target partition Analysis: Background rationale, problems and prospects, Psychometric Society, Iowa City.

Pruzek, R.M. (unpublished). Content Specification.

Pruzek, R.M., and Lehrer, R. (unpublished). A general method for structural analysis of relationships among qualitative variables. University at Albany, State University of New York.

Pruzek, R.M., Stegman, C.A., and Pfeiffer, R.A. (1972). On the Analysis of Partitioned Data, paper presented at the 1972 American Education Research Association Convention.

Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. Paper presented at the American Educational Research Association annual meeting, Chicago.

Rock, D.A., and Pollack-Ohls, J. (1987). Measuring Gains: A New Look at an Old Problem. Princeton, New Jersey: Research Statistics Group, Educational Testing Service.

Roid, G., and Haladyna, T. (1982). A Technology for Test Item Writing. New York: Academic Press.

Sato, T. (1975). The construction and interpretation of S-P tables. Tokyo: Meiji Tosho.

Scandura, J.M. (1977). Structural approach to instructional problems. American Psychologist 32: 33-53.

Scheuneman, J., and Steinhaus, K. (April, 1987). Some theory of item difficulty and discrimination. Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Stenner, A.J., Smith, M., and Burdick, D.S. (1983). Toward a theory of construct definition. Journal of Educational Measurement, 20, 305-316.

Sternberg, R.J. (1985). Beyond IQ: A Triarchic Theory of Intelligence. New York: Cambridge University Press.

Tatsuoka, K.K. (1983) Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20, 345-354.

Wiley, D.E. (1967). Latent Partition Analysis, Psychometrika, 32(2), 183-193.

Wilson, D., Wood., R.L., and Gibbons, R. (1984). TESTFACT: Test scoring and item factor analysis, User's Guide, Version 2.2. Chicago: Scientific Software.

Wilson, M.R. (1985) Measuring Stages of Growth: A Psychometric Model of Hierarchical Development. Occasional paper No. 19. Hawthorne, Australia: Australian Council for Educational Research.

———. (1984) A Psychometric Model of Hierarchical Development. Unpublished PhD thesis, University of Chicago, Department of Education.

PROFILE 1

CATEGORY	ITEM NUMBER
Work Registration Requirements	19
Categorical Determination	41, 40, 35, 24, 22, 14, 25, 17, 15, 12
Child Support Enforcement (Conflict)	36, 16
Food Stamp Eligibility	42, 37
Interviewing Skills	43, 34, 27, 21
Assessing the need for referrals	49, 44
Application process	29, 3, 2, 50, 47, 1, 6
Financial Eligibility for Public Assistance	
A) Budgeting of Income	48, 38, 32, 39, 31, 11, 9, 7, 5, 8, 4
B) Resources	18, 13
MA/Third Party Health Insurance	20, 28, 33
Non-compliance	10, 30
Verification requirements	23, 26, 46, 45

TABLE 1B: TARGET PARTITION ANALYSIS

CASE SUMMARY

| | | |

PROFILE 2

CATEGORY	ITEMS
Medicaid and Food Stamps Eligibility Criteria	
A. Food Stamps	42, 37
B. Medicaid	28, 33, 20
Documentation/Verification (tough)	26, 2, 3, 8, 10, 30, 23, 46, 45, 29
Resources	13, 18
3209	50, 47, 1, 6
Budgeting (easy)	4, 48, 38, 32, 39, 31, 11, 5, 9, 7
Interviewing	21, 26, 43, 34
Mandated and Non-mandated Referrals (tough)	19, 49, 44, 36, 16
Categorical Determination (easy)	35, 40, 22, 12, 41, 25, 17, 15, 24
Miscellaneous	14

TABLE 2A: TARGET PARTITION ANALYSIS

			22334	22 23 1 44	11	45	133	343	3242	11443	123122144	1	CATEGORY	ODD	FIT	ITEMS
	TARGET ORDER	-	08372	3639080256	83	6170	4189759182		4137	69496	525245701	4		2	8	
	TARGET PART.	-	11111	2222222222	33	4444	5555555555		6666	77777	888888888	9		2	5?	
CASE	#C	QST												7	19	
1	9	0.106	77733	9955868599	66	5555	6666666666		4444	21442	222222222	2		7	16, 36?	
2	5	0.251	33651	5566252525	11	2552	1111222111		5555	31555	542444544	4		4	35	
3	11	0.088	22244	9999999999	33	9999	5555575555		0000	16001	787877587	A				
4	7	0.111	37777	7777727777	37	5556	3333232333		1111	67666	445444444	4				
5	12	0.103	88811	2202853232	44	7777	4444545444		6666	05660	957599959	A				
6	6	0.148	11313	3211111131	33	4444	3555515555		2222	11113	664666666	1				

MEAN QST=.136, SD=.061

TABLE 2B: CONFUSION MATRIX OF CATEGORIES

Weighted diagonal mean=6.72
Weighted off-diagonal mean=1.07
Category clarity=6.28

	40.00								
	25.02	54.73							
	18.40	12.65	83.00						
	3.40	29.95	0.00	88.67					
	6.76	5.39	50.00	2.55	70.67				
	6.80	13.50	0.00	8.50	0.00	100.00			
	18.04	21.02	8.40	5.10	4.76	30.20	35.00		
	1.51	3.78	0.00	8.39	4.33	3.78	10.16	62.11	
	10.20	11.90	0.00	0.00	1.70	0.00	20.20	42.56	0.00
		*			**		*	**	
ITEMS/CATEGORY	5	10	2	4	10	4	5	9	1

* Difficult categories for this trainer to consolidate.
** Comparatively easy category to consolidate

PROFILE 3

CATEGORY	ITEMS
Food Stamps	37, 42
Application	2, 23, 26
Certification (toughest)	29, 46
Recertification (easy)	10, 45
30 + 1/3	11, 39
Budgeting Expenses	4, 31, 32, 38, 48
Resources (tough)	13, 18
ADC-U	12, 22, 40
Employment	19
Monthly reporting	7, 8, 9
Interviewing	21, 27, 34
Difficult Client	43
Referrals	44, 49
APP-IND-Authorization Change Form	1, 6, 35, 47, 50
Medicaid or Third Party Health Insurance	20, 28, 33
Category (easy)	15, 17, 24, 25, 41
IV-D	3, 16, 36
Miscellaneous	14
Miscellaneous	30

TABLE 3A: TARGET PARTITION ANALYSIS

CASE SUMMARY

			13|34|2 224|41|33 133 114| 4 121|442234| 4 35|232|42211|1|3	CATEGORY	OLD FIT ITEMS
TARGET ORDER -			366|72|92366|50|1258894138|9708922|347149|67150|038|14557|4|0	1	3?
TARGET PART. -			000|11|22222|33|4444444444|5555555|666666|77777|888|99999|A|B	2	45
CASE	#C	QST		2	13, 18?
1	9	0.104 -	500|33|55999|98|6666666666|6626122|444444|55525|777|22222|2|8	3	44, 49?
2	5	0.222 -	535|51|55555|22|1121111111|2245144|555555|25522|353|44455|4|2	4	35
3	11	0.127 -	911|44|99999|99|5573555535|5589688|000000|99979|222|77775|A|9	6	17?
4	7	0.134 -	766|77|77777|77|3333333373|2242744|161116|55555|377|44444|4|7		
5	6	0.190 -	113|13|11321|31|5513553535|5561166|212221|44444|131|66666|1|1		
6	9	0.109 -	277|11|22222|22|5553555535|5582788|676667|44484|111|88888|9|2		

MEAN DIFCE=0.47, SD=0.047

TABLE 3B: CONFUSION MATRIX OF CATEGORIES

Weighted diagonal mean=4.68
Weighted off-diagonal mean=1.03
Category clarity=4.54

	33.67											
	19.67	67.00										
	37.98	31.70	70.30									
	25.17	25.00	61.80	67.00								
	3.40	12.60	3.72	6.70	68.80							
	13.62	7.07	12.40	14.36	20.63	23.00						
	22.33	11.17	22.33	2.83	0.57	4.76	73.33					
	15.60	3.40	25.28	23.40	1.34	8.60	6.80	69.80				
	20.22	38.83	24.40	19.50	9.00	5.62	9.44	2.27	66.67			
	15.73	3.40	6.80	0.00	3.74	34.77	6.80	12.04	2.27	83.30		
	22.33	8.50	10.20	8.50	1.70	26.29	5.67	3.40	11.33	43.20	0.00	
	28.00	25.00	60.20	83.50	5.00	16.71	5.67	23.40	22.33	0.00	17.00	0.00
		*	**	*					**			
ITEMS/CATEGORY	3	2	5	2	2	10	7	6	5	5	1	1

* These categories were described as difficult to consolidate.
** These categories were described as easy to consolidate

PROFILE 4

CATEGORY	ITEMS
Eligibility (tough)	42, 13, 18, 4, 11, 31, 39, 48, 38, 32, 19
Maintaining	35, 50, 9, 7, 5, 10, 30, 45, 6
Benefits	16, 20, 28
Funding	14, 12, 40, 24, 41, 25, 22
Processing (easy)	8, 2, 1, 3, 15, 17, 23, 29, 47, 26, 46, 21, 27, 43, 34, 37, 33, 49, 36, 44

TABLE 4A: TARGET PARTITION ANALYSIS

CASE SUMMARY

| | | |

TABLE 5: SUMMARY STATISTICS FOR THE TARGET PARTITION ANALYSIS

Case No.	No. of Categories	Average Q_{st}	Category Clarity
3	5	0.252	3.82
4	6	0.202	4.75
5	7	0.170	6.38
1	9	0.161	4.84
2	9	0.136	6.28
6	11	0.163	5.94
7	12	0.147	4.54